Exhibit 1

```
1
 1
                UNITED STATES DISTRICT COURT
 2.
               SOUTHERN DISTRICT OF NEW YORK
 3
     In Re:
 4
    OPENAI, INC.,
    COPYRIGHT INFRINGEMENT
 5
    LITIGATION
                                       Case No.:
                                       I:25-md-03143-SHS
 6
    This Document Relates to:
 7
    Case No.: I :23-cv-08292-SHS
 8
    Case No.: I :23-cv-l 0211-SHS
     ----/
 9
10
11
12
       **HIGHLY CONFIDENTIAL - ATTORNEY'S EYES ONLY**
13
14
         VIDEO-RECORDED DEPOSITION OF MICHAEL TRINH
15
                 30(b)(6) CAPACITY FOR OPENAI
16
                   Friday, July 25, 2025
17
18
19
20
     Stenographically reported by:
21
     LORRIE L. MARCHANT, RMR, CRR, CCRR, CRC
     California CSR No. 10523
22
    Washington CSR No. 3318
23
    Oregon CSR No. 19-0458
    Texas CSR No. 11318
24
    Job No.: 2025-992367
25
    California Firm Registration No.: 179
```

BY MR. SMYSER:

- Q. Okay. Apart from that Slack message involving Mr. Radford describing his downloading of the dataset that became known as Books1, were there other documents that you recall reviewing in preparation for this deposition?
 - A. Yes.

- O. What are they?
- A. I recall reviewing source -- or, sorry, let me -- before I move on from Slack, I recall reviewing other Slack messages discussing Books2.
- Q. Okay. Who were those Slack messages from or involving?
 - A. I believe they all involved a former employee named Ben Mann.
 - Q. And what was the content of those messages?
 - A. His general discussions about his review of Booksl and -- or, sorry, the dataset that is identified here as Booksl and his efforts to create a -- a second dataset that is known under these object- -- or responses and objections as Books2.
 - Q. And what did the messages say about his review of Books1?
 - A. He discussed that he wanted to down- -- or he -- he would create a second one.

- Q. A second Books dataset; is that what you understood?
 - A. Correct.

- Q. And so what did he say about his efforts to create Books2?
- A. He describes -- he describes his collections of -- of the data files that be- -- about the time that he did.
 - Q. And where did he collect the files from?
- 10 A. So I also reviewed excerpts of -- from his
 11 deposition in which he discusses using a technology
 12 called "torrenting."
- T-O-R-R-E-N-T-I-N-G-S. Sorry.
 - Q. And so he used a technology called torrenting, and what did he torrent?
 - A. I believe he torrented files also associated with that website known as Library Genesis.
 - Q. And so apart from these Slack messages regarding Mr. Radford and Mr. Mann, were there other Slack messages you reviewed in preparation for this deposition?
 - A. Those are the ones that come to mind.
- Q. Do you recall any other participants in either of those Slack messages apart from

45 Same objections. 1 MR. GOLDBERG: Objection 2 to the form. Beyond the scope. 3 THE WITNESS: I'm prepared to testify about 4 these versions. 5 BY MR. SMYSER: That wasn't my question, sir, with respect. 6 Ο. Are you aware of any other versions of the 7 Books1 and Books2 datasets apart from the four we 8 discussed and the intermediate copies you just 9 10 mentioned? MR. GOLDBERG: Objection to the form. 11 12 Beyond the scope. Asked and answered. 13 THE WITNESS: I am aware that we -- that 14 there was a copy of one of these datasets recovered, 15 and I don't know if that's a -- if that's a version 16 or not. BY MR. SMYSER: 17 Do you know if that's the copy that's been 18 Ο. 19 produced to us in this litigation? 20 That is my understanding. Α. 21 Okay. Apart from the copy that you Ο. 22 understand was produced to us in litigation, the 23 intermediate copies that we've discussed, and the 24 four versions that we've discussed, are you aware of 25 any versions of the Books1 and Books2 datasets?

58 1 The -- the decision to delete THE WITNESS: 2 was a company decision and so there are -- that --3 it's a company decision. 4 BY MR. SMYSER: 5 Okay. Let's talk about some of the others Ο. that don't have asterixis. 6 7 Che Chang, why was -- I'm sorry, Mr. or Mrs. Chang, why were they involved? 8 Mr. and it's Che. 9 Α. 10 Q. Apologies. Yeah. He's -- Mr. Chang is a lawyer. 11 Α. 12 And why was he involved? Q. Okay. There were -- these -- these discussions 13 Α. 14 were privileged; but, generally speaking, they 15 involved legal -- legal strategy, copyright risk, 16 and product legal strategy. Mr. Chang is a product -- well, was then a 17 product counsel, among many things, but product 18 19 counsel was his -- is his primary function then. Okay. And Mr. Kwon, why was he involved? 20 Ο. 21 Mr. Kwon was the general counsel at the Α. 22 time. 23 Okay. So returning to Mr. Pachocki, who's Q. 24 not an attorney; correct? 25 Α. Correct.

59 What were the nonprivileged facts that he 1 Q. 2 conveyed during these oral communications? 3 MR. GOLDBERG: Objection. Beyond the scope 4 and assumes facts. 5 And, again, as we've meet -- met and conferred about, as you've testified, these oral 6 7 communications were privileged. MR. SMYSER: Are you instructing him not to 8 9 answer or ... 10 MR. GOLDBERG: There's no predicate on which I -- I'd like to know whether he knows an 11 answer. But if -- but if the answer is based on 12 13 privilege, I'm instructing him not to answer. if the answer is "I don't know," he can answer that 14 15 because that's not privileged. BY MR. SMYSER: 16 17 Okay. Do you know if Mr. Pachocki conveyed Ο. any nonprivileged or technical facts during these 18 19 oral communications you're discussing? 20 MR. GOLDBERG: Same objection. Beyond the 21 scope. 22 THE WITNESS: I was not informed of any. 23 BY MR. SMYSER: 24 Q. Any?

Α.

25

My understand- -- what I learned was that

- the discussion was privileged, and it discussed -well, as I said -- as I mentioned before -- legal
 strategy, copyright, legal questions, and product
 legal decisions.
 - Q. So you are not aware of any nonprivileged facts conveyed during those oral communications?
 - A. I -- this discussion -- the discussion to delete was led by legal, and it was -- and it -- it incorporated those three things that I -- I mentioned.
- Q. I understand that.
- But legal brought in these technical
 people, correct, Bob McGrew, Jacob Pachocki, Alex
 Paino, Lilian Weng, and Jeff Wu; right?
- 15 A. Correct.

1

2

3

4

5

6

7

8

9

10

11

16

17

19

20

- Q. And brought them in due to their technical role at OpenAI; correct?
- 18 A. Yes.
 - Q. And so what technical knowledge did they have that was useful to this process? If you know.
 - A. I don't know the specifics.
- Q. Are you aware of any specifics, any specific technical knowledge that they had that explains why they were consulted as part of this series of oral communications?

1 models that they may have been used to train at the 2 time of these oral communications we're discussing? 3 Objection to the form. MR. GOLDBERG: 4 THE WITNESS: I am aware that these -- the 5 datasets were no longer being used generally. BY MR. SMYSER: 6 What do you mean by "generally"? 7 Okav. Ο. So, for example, if they were already 8 Α. incorporated in a training run and had already been 9 10 used, those models still -- were still -- still in existence and still, you know, operating. 11 12 Q. And you're not aware of the datasets being 13 used for anything else at the time of these oral 14 discussions? 15 Α. Correct. Okay. When did those oral discussions take 16 Ο. 17 place? June into July 2022. 18 Α. 19 Ο. Do you know specific dates? I believe there was -- a lot of activity 20 Α. 21 was in -- particularly in mid-July, 16th or 17th. 22 And do you know the earliest date in June 0. 23 that the discussions took place? 24 Not specifically, no. Α. 25 Do you know generally in June when the Q.

65 Books1 and Books2 used to train GPT-3 deleted? 1 2 Α. Yes. And were the versions of Books1 and Books2 3 Ο. 4 used to train GPT-3.5 deleted? 5 Α. Yes. And all of the versions used for those two 6 Ο. models were deleted; is that correct? 7 MR. GOLDBERG: Objection to the form. 8 9 THE WITNESS: We did recover a copy of 10 one -- of a set of versions. BY MR. SMYSER: 11 12 Understood you recovered a copy. Q. 13 But were all the versions deleted 14 initially? 15 MR. GOLDBERG: Objection to form. 16 THE WITNESS: That was the intent. 17 BY MR. SMYSER: Was it the actuality? 18 Ο. 19 Α. Like I said, we recovered that one copy 20 of -- of -- we -- we recovered one copy. 2.1 Were all the copies used to train GPT-3 Ο. 22 deleted prior to any recovery? 23 MR. GOLDBERG: Objection to the form. Asked and answered. 24 25 THE WITNESS: We -- like I said, we

66 1 recovered a copy. It was not -- we found a copy. It was not -- I'm -- I'm pausing when you say the --2 when I use the word "recovered," I didn't mean, 3 4 like, we undeleted it or something. Like, we -- we 5 located a copy of what we believed to be the version that trained Books3. 6 BY MR. SMYSER: 7 Okay. So for the versions used to train 8 Ο. Books3 --9 10 Sorry. Sorry. For the versions -- I'm Α. tying myself in a knot. 11 For the versions of Books1 and -2 used to 12 13 train GPT-3, we recovered a copy of what we believe 14 to be those datasets, that version. 15 0. So for the versions used to train GPT-3, 16 not all of those versions -- or rather, not all of 17 those copies were deleted in 2022; correct? Because we know of that one copy, we know 18 19 at least one copy, like, was -- was not deleted. 20 Ο. Okay. Where was that copy? 21 We -- OpenAI recovered a copy of those 22 versions in -- it was -- my understanding, it was in 23 a -- it was somewhere in a (inaudible) --24 (Stenographer interrupted for clarification 25 of the record.)

67 1 THE WITNESS: Sorry. Azure Blob store. 2 BY MR. SMYSER: 3 And do you know why it was not deleted Ο. 4 along with the other copies? 5 MR. GOLDBERG: Objection. I think that's 6 beyond the scope. 7 THE WITNESS: No one was aware that that copy existed. 8 9 BY MR. SMYSER: 10 Q. Okay. Do you know where on Azure Blob store it was? 11 12 I don't have that specific location, no. Α. 13 Do you have a general understanding of Q. 14 where it was on Azure Blob store? 15 Α. It was -- my understanding is it was not in -- it was in a section of Blob store that was not 16 widely available -- widely availably -- sorry, 17 widely available, and it wasn't part of the areas 18 19 that we commonly store datasets. 20 Okay. Is there any other information you Ο. 21 can give me about what this section of the Blob 22 store was? We searched broadly, and they -- Ian Sohl 23 Α. located this -- this -- these files. 24 25 Would he know the specifics about where the Q.

25

Objection to the form.

the name of the dataset or something else?

MR. GOLDBERG:

69 I believe he searched across 1 THE WITNESS: 2 file names using strings. 3 BY MR. SMYSER: 4 So it's your understanding that he located Ο. 5 the recovered version of the Books1 and Books2 datasets used to train GPT-3 by searching for the 6 7 file names in Azure Blob store; correct? MR. GOLDBERG: Objection to the form. 8 That is my understanding, 9 THE WITNESS: 10 yes. BY MR. SMYSER: 11 12 And when he performed that same process Q. with respect to the names on Exhibit 279 for the 13 14 GPT-3.5 Books1 and Books2 datasets, he did not 15 locate copies of those; correct? 16 Α. He was only able to locate copy -- a copy of the versions that we believe were used to train 17 GPT-3.18 And the section of the Blob store that he 19 Ο. located these versions in, you mentioned it was not 20 2.1 widely available; is that right? 22 That is my understanding. Α. 23 Why was it not widely available? Q. 24 Beyond the scope. MR. GOLDBERG: 25 I don't know the specifics. THE WITNESS:

70 My general understanding is that this was a -- this 1 2 area -- this was not a -- to be honest with you, it 3 was old and forgotten. 4 BY MR. SMYSER: 5 Okay. Do you have any more information Ο. 6 about that section of the Blob store that you can 7 give me? Not in specifics, no. 8 Α. Any general information you can give me in 9 Q. 10 addition to what we've already discussed? I've described that it's old, and it was 11 Α. 12 not -- you know, not -- it was not the area where we 13 actually store datasets for use in training. 14 Q. Was it somebody's personal account? 15 Α. I -- I don't know. 16 So you don't know one way or another Q. Okay. 17 whether it was somebody's personal account? 18 Α. Correct. 19 Q. Okay. But Mr. Sohl would know? 20 Objection to the form. MR. GOLDBERG: 2.1 THE WITNESS: You could ask him. 22 BY MR. SMYSER: 23 Okay. Do you know if -- well, actually, Q. 24 let's step back. 25 Do you know how Alec Radford obtained the

30(b)(6), Attorneys Eyes Only 71 1 LibGen-1 dataset? 2 My understanding is that he accessed it via 3 direct download from a -- from a website. 4 Ο. Do you know what website? 5 Α. I believe it was associated with LibGen. And when you say "direct download," what do 6 Ο. 7 you mean by that? An HTTP GET request. 8 Α. And was that -- is that different from 9 Q. 10 torrenting, in your understanding? 11 Α. Yes. 12 How is it different from torrenting, in Q. 13 your understanding? 14 So an HTTP GET request -- sorry, HTTP GET Α. 15 request is a protocol -- or is a request under the 16 HTTP protocol, which is how you pull files from generally the Internet, commonly from a website. 17 is -- involves a client making a request directly to 18 19 a server. 20 And then -- and then the server responds to the GET request by providing the file or whatever is 21 22 requested by the GET request. 23 Now, a torrent is not part of the HTTP 24 protocol. It is -- refers generally to a

25

peer-to-peer -- my understanding is a peer-to-peer

72 1 model where the -- the information can come from 2 computers, besides the server, that you make the 3 request of. 4 Does OpenAI still have a copy of the HTTP Ο. 5 GET request? 6 MR. GOLDBERG: Objection. Beyond the 7 scope. THE WITNESS: HTTP GET requests are by 8 9 their nature ephemeral; right? It's, you know, 10 something that your computer issue- -- like, for example, if you go to a particular website, your 11 12 computer issues a GET request for the web page at 13 that website and then it gets a response. That's 14 just kind of normal network traffic. 15 So -- so the answer is, no. BY MR. SMYSER: 16 17 Does OpenAI have any records of that Ο. download? 18 19 MR. GOLDBERG: Objection to the form. Beyond the scope. 20 21 THE WITNESS: I did not review any. 22 BY MR. SMYSER: 23 And you're not aware of any apart from Q. 24 those that you didn't review? 25 MR. GOLDBERG: Same objection.

78 1 torrenting file from LibGen or not; correct? 2 Α. Correct. Okay. When were the versions of Booksl and 3 Ο. 4 Books2, apart from the recovered version, deleted? 5 Α. July 2022. Do you know approximately what days or day? 6 Ο. I think it's in that mid-July time frame. 7 Α. And at that time OpenAI and other AI 8 Ο. developers, they faced substantial legal uncertainty 9 10 about whether training their models on copyrighted works was fair use; is that correct? 11 12 MR. GOLDBERG: Objection to the form. 13 Beyond the scope. Incomplete hypothetical. Assumes 14 facts. 15 THE WITNESS: Could you repeat the 16 question? BY MR. SMYSER: 17 So in July 2022, OpenAI and other AI 18 Ο. 19 developers faced substantial legal uncertainty about 20 whether training their models on copyrighted works 2.1 was fair use; is that correct? 22 MR. GOLDBERG: Objection. Form. Beyond 23 Incomplete hypothetical. the scope. Assumes facts. 24 THE WITNESS: The -- the company hadn't 25 been sued, and that was probably almost a year

nonprivileged facts identifying the 1 2 persons directly involved in deletion of 3 Books1" -- apologies -- "in the deletion 4 of the Books1 and Book2 datasets and the 5 nature of the involvement of each person directly involved in the deletion of the 6 Books1 and Books2 datasets." 7 Do you see that? 8 9 Α. I do. 10 And are you prepared to testify as to that Q. 11 scope? 12 Α. I am. 13 So who was directly involved in the Q. Okay. 14 deletion of the Books1 and Books2 datasets? 15 Α. I spoke with Alex Paino, and he's the one 16 who deleted the datasets. 17 Ο. Okav. So was he the only person directly involved in the deletion? 18 19 Α. He was the one who deleted it. There -- he 20 told me that Jeff Wu was involved in modifying source code that referred to the -- to the deleted 21 22 datasets. 23 Okay. And when you say "modifying source Q. code that referred to the deleted datasets," what do 24 25 you mean by that?

- A. With the datasets no longer there, soft -- or source code that referred to it and would look -- look for those datasets, they were modified to remove the references to the datasets.
- Q. And is it your understanding that the source code prior to those modifications is still within OpenAI's possession?
- A. Those would be -- the earlier versions

 pre -- pre-Jeff Wu's modifications still exist, yes.
- Q. Okay. And do you know if they've been produced in this action?
- 12 A. That is my understanding.

1

2

3

4

5

6

7

8

9

10

11

13

14

15

16

17

18

19

- Q. Okay. So by "directly involved," you mean either, in the case of Mr. Paino, that they deleted the datasets themselves or, in the case of Mr. Wu, that they modified source code that referred to those datasets; is that correct?
 - A. That is my understanding, yes.
- Q. Is there anything else meant by "directly involved"?
- 21 A. I don't imply anything else there with that 22 term.
- Q. Okay. When we were looking at Exhibit 279 earlier, the third table here, you recall there were a number of technical people listed who were not

- Mr. Wu or Mr. Paino; correct?
- 2 Α. Correct.

1

3

4

5

6

7

8

9

10

11

12

13

15

16

17

18

- And it's your testimony that those Ο. technical people were not directly involved in the deletion of Books1 or Books2; is that correct?
 - I believe they were involved in the discussions, but in terms of the -- the operation of deletion, it's -- I discussed Mr. Paino and Mr. Wu.
 - So just to be clear, when you're Q. Okay. saying they were directly involved in the deletion, you mean directly involved in effectuating that deletion?
- Α. The technical, like, deletion, yes.
- 14 Q. Okay. And then apart from Mr. Paino and Mr. Wu, was there anyone else, to your knowledge, directly involved with the technical aspects of the deletion?
 - I did not find -- I -- I did not find any information beyond those two.
- Okay. And so the other individuals listed 20 Ο. on Exhibit 279 you indicated are directly involved 21 with the decision to delete Books1 and Books2; 22 23 correct?
- 24 MR. GOLDBERG: Misstates testimony.
- 25 THE WITNESS: They were involved in the

yes.

1

2

3

4

5

6

7

8

9

10

11

16

17

18

19

20

2.1

- Q. And the one that Tom Rubin worked at, at the time, was one of those law firms providing legal advice in connection with the decision to delete Books1 and Books2; correct?
 - A. Correct.
- Q. And what about Joe Gratz? Was he involved in the decision to delete Books1 and Books2?
- A. He was -- I -- he was employed at

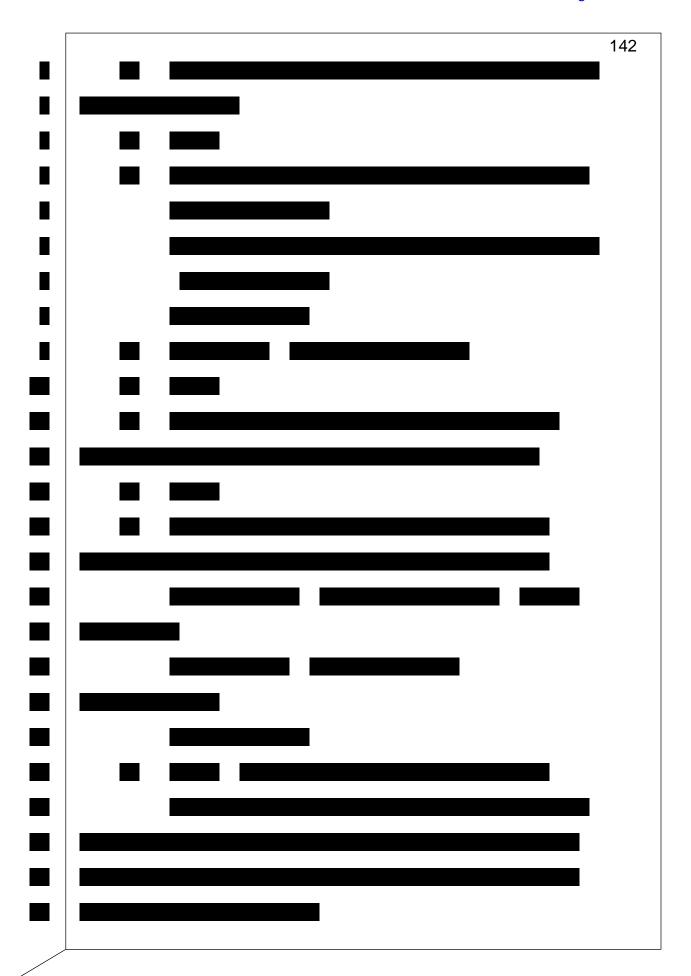
 Morrison & Foerster, which is another law firm that

 gave advice to OpenAI, Jason and Che.
- 12 Q. In connection with the deletion of Books1
 13 and Books2?
- 14 A. Mo -- MoFo, Morrison & Foerster -- I'll say
 15 that on the record -- gave advice on that, yes.
 - Q. Okay. And do you know for sure one way or another if Mr. Gratz was involved in that decision?
 - A. I -- I don't have any information on that.
 - Q. Okay. And what about Tom Gorman, was he involved in the decision to delete Books1 and Books2?
- A. I believe he's employed at Keker Van Nest, and I don't have any information that Keker & Van Nest provided guidance on -- on the -- on this issue.

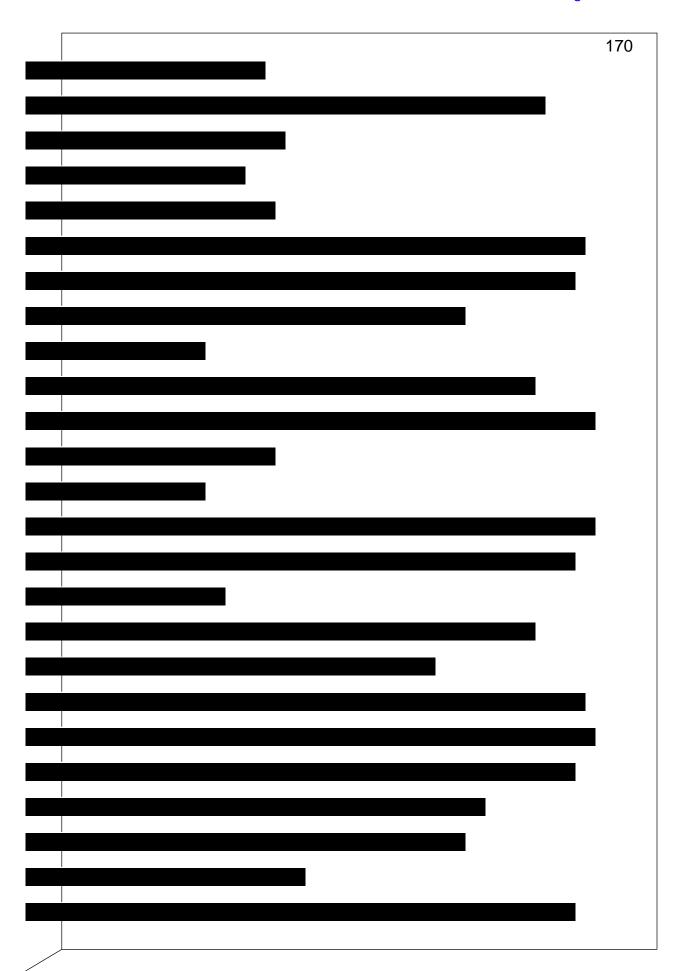
126 1 page 14 of Exhibit 278. 2 Thank you for the page number. So do you see the first sentence of the 3 Ο. 4 first full paragraph on that page (as read): 5 "Without waiving any of these objections, OpenAI will provide 6 7 nonprivileged facts regarding the technical steps undertaken for the 8 deletion of the Books1 and Books2 9 10 datasets"? 11 Α. Yes. 12 And you're prepared to testify as to those? Q. 13 Yes. Α. 14 Okay. So did OpenAI undertake technical Q. 15 steps to identify the copies of the Books1 and Books2 datasets that needed to be deleted? 16 17 Α. Yes. What were those technical steps? 18 O. 19 Α. Alex Paino is the engineer who worked with those datasets and identified locations in the --20 2.1 where -- the locations -- datasets that had been 22 used to train GPT-3 and GPT-3.5. 23 And what technical steps did he take to Q. 24 perform that identification? 25 He was aware of their -- their use in Α.

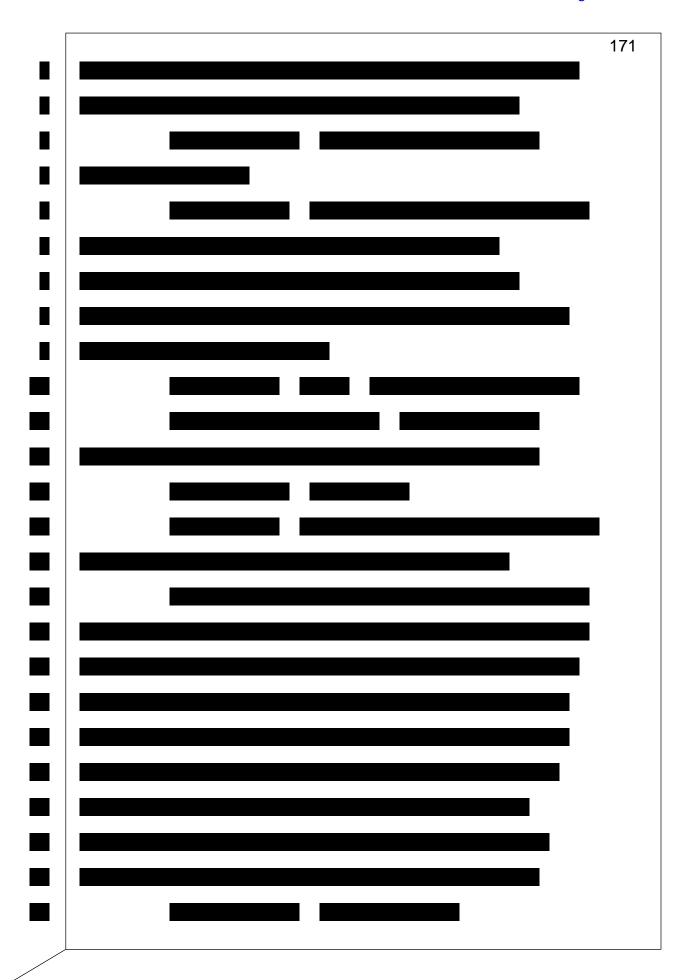
- 1 training and the steps that had been taken in
- 2 | between the GPT-3.5 versions and the -- and the
- 3 | GPT-3.5 versions and identified those locations and
- 4 deleted the files there.
- 5 Q. And when you say he identified the
- 6 | locations, were those on, as you've called it, the
- 7 | Azure Blob Storage?
- 8 A. Yes.
- 9 Q. Was that the only place they were?
- 10 A. That is the location of -- that is the --
- 11 | the location that was used to train, yes.
- 12 Q. Okay. And just to be clear, you mentioned
- 13 | the intermediate copies as part of that -- the
- 14 | penultimate answer; correct?
- 15 A. Yes.
- 16 O. And were those also on the Blob Storage, is
- 17 | what I'm trying to get at, to your knowledge?
- 18 A. I -- Blob Storage is the place that we
- 19 | would -- that we use generally for these -- this
- 20 | sort of thing. So there may have been processing
- 21 | elsewhere, but that's where the -- all -- that's
- 22 where the datasets are stored.
- Q. Okay. And I think you mentioned that
- 24 | Alex Paino made this identification by virtue of the
- 25 | fact that he worked with these datasets previously;

128 1 is that correct? 2 Α. Yes. And so did he perform any, I quess, 3 Ο. 4 additional technical steps apart from his own 5 knowledge that helped him identify where these 6 copies were? He -- he was the person at the company who 7 Α. knew where these datasets were and how they were 8 used so he -- he was instructed to delete. 9 10 Q. Understood. I'm trying to understand how he figured 11 12 out, "Here's every copy of this I need to delete." 13 And is the testimony just that he -- he 14 knew where they were because he previously worked 15 with them, or did he also do some technical search 16 to determine where the other copies were? 17 Objection to the form. MR. GOLDBERG: I believe he did a technical 18 THE WITNESS: 19 And he's -- he's familiar with these 20 datasets. He's the most -- he -- and he was -- you 21 know, I think the fact that we were unable to 22 recover any other copies -- well, except for the one 23 I previously described -- you know, kind of -- I think that speaks for itself. 24 25 ///



143 1 Α. Correct. And it was Mr. -- was it Mr. Sohl or 2 Ο. 3 Mr. Paino who located it? 4 Α. Sohl. 5 Mr. Sohl located it. Ο. 6 Α. Yeah. Do you know when he located it? 7 Ο. I think it was sometime last year. 8 Α. So in 2024 sometime? 9 Q. 10 Α. I believe so, yes. And how does OpenAI know that the versions 11 Q. 12 recovered by Mr. Sohl correspond to the versions used to train GPT-3? 13 14 The files that he recovered involve Α. 15 something called a UUID. And we've seen evidence that it's the same UUID that was used in the GPT-3 16 17 training run. What is a UUID? 18 Ο. Okay. 19 Α. It a technical acronym, and I actually --20 it's something about a universal identifier. is -- this is an external one's -- external 2.1 22 definition. 23 So it's commonly used. It's not specific But it is a unique identifier that you 24 to OpenAI. 25 can -- that you can create. I believe the





STENOGRAPHER'S CERTIFICATE

I, LORRIE L. MARCHANT, Certified Shorthand Reporter, Certificate No. 10523, for the State of California, hereby certify that MICHAEL TRINH was by me duly sworn/affirmed to testify to the truth, the whole truth and nothing but the truth in the within-entitled cause; that said deposition was taken at the time and place herein named; that the deposition is a true record of the witness's testimony as reported to the best of my ability by me, a duly certified shorthand reporter and a disinterested person, and was thereafter transcribed under my direction into typewriting by computer; that request [] was [X] was not made to read and correct said deposition.

I further certify that I am not interested in the outcome of said action, nor connected with, nor related to any of the parties in said action, nor to their respective counsel.

IN WITNESS WHEREOF, I have hereunto set my hand this 6th day of August, 2025.

LORRIE I. MARCHANT RMR CRR

LORRIE L. MARCHANT, RMR, CRR, CCRR, CRC Stenographic Certified Shorthand Reporter #10523